

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 9, 2003, 08:33:19 ; Search time 819 seconds
(without alignments)
11317.735 Million cell updates/sec

Title: US-09-601-965C-1
Perfect score: 4116
Sequence: 1 aaaaaaaaaaagctccccc.....tattatgattcattcattc 4116

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N_Geneseq_101002.*
1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	4116	100.0	4116	20	AAH86690 DNA encoding a TAF
2	91	2.2	3399	17	AAT05868 Chicken leucocytos
3	88.6	2.2	1998	21	AAH70212 Plasmodium falcipar
4	86.6	2.1	3579	21	AAH70099 Plasmodium falcipar
5	84	2.0	4056	21	AAH70225 Plasmodium falcipar
6	82.2	2.0	1686	16	AAH07587 DNA encoding leuco
7	79.8	1.9	1080	24	AB199537 Mouse ischaemic co
8	79.2	1.9	49999	20	AAZ23895 Murine LOB8 homolo
9	78.4	1.9	400	23	AAH75460 DNA encoding novel

10	78.4	1.9	963	23	AAH68580 DNA encoding novel
11	78.2	1.9	801	23	AAH90729 DNA encoding novel
12	77.6	1.9	1083	23	AAH6745 DNA encoding novel
13	77	1.9	372	23	AAH75452 DNA encoding novel
14	77	1.9	1072	23	AAH90738 DNA encoding novel
15	76.8	1.9	700	20	AAH8907 EST clone IN200.
16	76.8	1.9	1391	21	AAH27866 Protein regulating
17	76.8	1.9	2554	21	AAH92314 Human cyclin 30 p
18	76	1.8	5994	21	AAH70222 Plasmodium falcipar
19	76	1.8	6621	21	AAH70188 Plasmodium falcipar
20	75.6	1.8	4167	22	AAH8224 Human ovarian and
21	75.6	1.8	4167	22	AAH29236 Genomic sequence #
22	75.6	1.8	4167	22	AAH30018 Human lung antigen
23	75.6	1.8	4167	22	AAH05022 Human reproductive
24	75.6	1.8	4167	22	AAH07543 Human reproductive
25	75.6	1.8	4167	22	AAH28711 Genomic sequence #
26	75.6	1.8	4167	22	AAH68040 Human immune/haema
27	75.6	1.8	4167	22	AAH9382 Human digestive sy
28	75.6	1.8	4167	23	AAH97915 Human testicular a
29	75.6	1.8	6668	24	ABH3697 Human immune syste
30	75.2	1.8	507	13	AAH32658 LEMMI 10 cDNA. Ly
31	75.2	1.8	696	22	AAH84231 Human cell death p
32	75.2	1.8	699	22	AAH84230 Human cell death p
33	75.2	1.8	717	22	AAH84229 Human cell death p
34	75.2	1.8	774	22	AAH84228 Human cell death p
35	75.2	1.8	819	22	AAH84227 Human cell death p
36	75.2	1.8	1669	22	AAH84226 DNA encoding novel
37	75.2	1.8	2427	23	AAH69554 DNA encoding novel
38	75.2	1.8	2427	23	AAH88862 Human immune syste
39	75.2	1.8	14006	24	ABH33958 Drosophila melanog
40	75	1.8	37996	23	AAH07876 DNA encoding novel
41	74.8	1.8	510	23	AAH69539 DNA encoding novel
42	74.8	1.8	510	23	AAH71141 DNA encoding novel
43	74.8	1.8	510	23	AAH90687 DNA encoding novel
44	74.2	1.8	1527	21	AAH70121 Plasmodium falcipar
45	74.2	1.8	3738	21	AAH70178 Plasmodium falcipar

ALIGNMENTS

RESULT 1	AAH86690	AAH86690 standard; DNA; 4116 BP.
ID	AAH86690	
AC	AAH86690	
XX		
DT	14-OCT-1999	(first entry)
XX		
DE	DNA encoding a TAF-145 protein of Candida albicans.	
XX		
KM	TATA-box binding protein-associated factor; TAF-145; gene transcription;	
KW	fungus gene transcription; fungal infection; ss.	
XX		
OS	Candida albicans.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	330..3812
FT		/*tag= a
XX		/transl_except= (pos: 3267..3269, aa: Ser)
PN	W09940199-A2.	
PD	12-AUG-1999.	
XX		
PF	08-FEB-1999;	99WO-US02940.
XX		
PR	09-FEB-1998;	98US-0024100.
XX		
PA	(SCRI-) SCRIPTGEN PHARM INC.	
XX		
PI	Long F, Thompson CM, Wobbe RC;	
XX		

606741100

Db 3961 CTGAGTTGTCMAAATTAAGATGGAACACGCAATGCATTTTGTCTCATCATCAACAAACGACATGA 4020
QY ACGAGAGTGTGATGATGATTTATGTCAGTAATATAGTTTACATTTTAAATGACATA 4080
Db 4021 ACGAGAGTGTGATGATGATTTATGTCAGTAATATAGTTTACATTTTAAATGACATA 4080
QY 4081 TAAACAATGATGATGATTTATGTCAGTAATATAGTTTACATTTTAAATGAT 4116
Db 4081 TAAACAATGATGATTTATGTCAGTAATATAGTTTACATTTTAAATGAT 4116

RESULT 2

AA05868
AA05868 standard; DNA; 3399 BP.

AC AA05868;

XX 14-AUG-1996 (first entry)

DE Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.

XX Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
immunisation; vaccination; ss.

XX Chicken leucocytozoan.

OS Location/Qualifiers

XX Key 1..3399

FT CDS /tag= a

FT misc_feature 1150..3218

FT /tag= b

FT /note= "Fragment referred to in the claims, for use as insert in a recombinant vaccine against chicken leucocytozoan disease"

XX JP07284392-A.

XX 31-OCT-1995.

XX 19-APR-1994; 94JP-0080643.

XX 19-APR-1994; 94JP-0080643.

XX (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.

XX (KITA) KITASATO KENKYUSHO SH.

XX MPI; 1996-006311/01.

XX P-PSDB; AAR97866.

XX Chicken leucocytozoan immunogenic protein - used in a recombinant vaccine against chicken leucocytozoan disease

XX Claim 6; Page 6-9; 35pp; Japanese.

XX AA05868 encodes a chicken leucocytozoan immunogenic protein, this DNA or a fragment of it can be used in a recombinant vaccine to immunise against chicken leucocytozoan disease. The DNA is used in a vector and operatively linked to an expression regulatory sequence as in standard practice.

XX Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other;

QY Query Match 2.28; Score 91; DB 17; Length 3399;

Best Local Similarity 43.7%; Freq. No. 1.6e-08; Matches 458; Conservative 0; Mismatches 585; Indels 6; Gaps 1;

QY 2421 ATGCAAAATGACAAAGATTGAAATTTATGAAATACCAAGACAGTGAGACCA 2480

Db 1729 ATACATGAG 1788

QY 2481 GGGTATTGAAAG 2540

Db 1789 GAGTATATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1848

QY 2541 ATGATTAACACCAAGATTCTTCAATGATGACACATATGCAATTTTGACACAGACTTA 2600
Db 1849 CATGAG 1908
QY 2601 GATGATTAATATGATGATTTATGATGATGATGATGATGATGATGATGATGAT 2660
Db 1909 GAG 1968
QY 2661 GGTGATTAAG 2720
Db 1969 GAG 2028
QY 2721 GATTAAG 2780
Db 2029 GAG 2088

QY 2781 GATTAAG 2840
Db 2089 CATGAG 2148

QY 2841 TCAAG 2900
Db 2149 GTTACATGAG 2208

QY 2901 AGAATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2960
Db 2209 GAG 2268

QY 2961 CCAATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3020
Db 2269 GTATATCATGAG 2328

QY 3021 CCGTATTTTACCCCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3080
Db 2329 CATGAG 2388

QY 3081 AGTTTACAG 3140
Db 2389 GAG 2448

QY 3141 CATGAG 3199
Db 2449 GAG 2508

QY 3200 -----TCATGATTTGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3254
Db 2509 GTTACATGAG 2568

QY 3255 CAACAGATCTCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3314
Db 2569 GAG 2628

QY 3315 GTTTGAGATTTACTGAG 3374
Db 2629 GAG 2688

QY 3375 TTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 3434
Db 2689 GTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 2748

QY 3435 GAGTATTTGAG 3493
Db 2749 GAG 2777

RESULT 3

AAA70212 standard; DNA; 1998 BP.

XX AAA70212;

XX 07-NOV-2000 (first entry)

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 08:34:49 ; Search time 10434 Seconds

(without alignments)
11480.463 Million cell updates/sec

Title: US-09-601-965C-1

Perfect score: 4116

Sequence: 1 aaaaaaaaaaagctccccc.....tattatgcattctcattgtc 4116

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_da.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_ets.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_da.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_ets.*

28: em_un.*

29: em_vl.*

30: em_hcg_hum.*

31: em_hcg_inv.*

32: em_hcg_other.*

33: em_hcg_mus.*

34: em_hcg_pln.*

35: em_hcg_rnd.*

36: em_hcg_mam.*

37: em_hcg_vrt.*

38: em_by.*

39: em_hcg_hum.*

40: em_hcg_mus.*

41: em_hcg_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1767.2	42.9	1794	6 AX489468	AX489468 Sequence
2	296.8	7.2	3950	8 SCYGR274C	Z73059 S. cerevisiae
3	296.8	7.2	7077	8 SCUI4954	U14954 Saccharomyc
4	296.8	7.2	8061	8 SCYCCG1	X84098 S. cerevisiae
5	275.8	6.7	36162	8 SPAC2G11	Z54354 S. pombe chr
6	125.6	3.1	811	11 CNS06EWE	AL39562 T3 end of
7	114.2	2.8	7218	6 166494	166494 Sequence 14
8	108.4	2.6	110000	2 PMA14P1_1	Continuation (2 of
9	105.4	2.6	115758	9 AC104634	AC104634 Homo sapi
10	103.4	2.5	43993	2 AC116965	AC116965 Dictyoste
11	99.2	2.4	62352	2 AC116990	AC116990 Dictyoste
12	99.2	2.4	224236	2 AL806518	AL806518 Mus muscu
13	98.2	2.4	309587	2 AL713990	AL713990 Mus muscu
14	97.4	2.4	59004	2 AC115680	AC115680 Dictyoste
15	96.6	2.3	218751	2 AL732595	AL732595 Mus muscu
16	96.6	2.3	259744	2 AL844181	AL844181 Mus muscu
17	96.2	2.3	230749	2 AC124415	AC124415 Mus muscu
18	95.8	2.3	155204	2 AC007926	AC007926 Trypanoso
19	95.2	2.3	155682	2 AC131189	AC131189 Mus muscu
20	94.8	2.3	160759	2 AC117082	AC117082 Dictyoste
21	94.4	2.3	111554	8 AP002460	AP002460 Arabidops
22	94.4	2.3	59869	2 AL844592	AL844592 Mus muscu
23	93.8	2.3	199285	2 AC103396	AC103396 Mus muscu
24	93.4	2.3	64693	2 AC128425	AC128425 Rattus no
25	92.8	2.3	163860	2 AC021628	AC021628 Mus muscu
26	92.8	2.3	184364	2 AC128425	AC128425 Rattus no
27	92.6	2.2	130118	2 AC122207	AC122207 Mus muscu
28	92.2	2.2	222585	2 AC101391	AC101391 Mus muscu
29	92.2	2.2	4995	3 PFAMESA	PFAMESA Plasmodium
30	92.2	2.2	247175	2 AC129593	AC129593 Mus muscu
31	91.8	2.2	158255	2 AC112577	AC112577 Rattus no
32	91.2	2.2	631	3 PPRPK51	X53026 P. falciparu
33	91.1	2.2	2069	6 E10125	E10125 DNA encodin
34	91.1	2.2	3399	6 E10126	E10126 DNA encodin
35	90.8	2.2	207750	2 AC109224	AC109224 Mus muscu
36	90.8	2.2	120044	2 AC129042	AC129042 Rattus no
37	90.6	2.2	153477	2 AC006278	AC006278 Plasmodiu
38	90.6	2.2	170032	2 AC116673	AC116673 Mus muscu
39	90.4	2.2	128921	2 AC116671	AC116671 Mus muscu
40	90.4	2.2	131310	2 AC114893	AC114893 Felis cat
41	90.2	2.2	81019	8 AB073159	AB073159 Arabidops
42	90.2	2.2	131310	2 AC114893	AC114893 Felis cat
43	90.0	2.2	143716	2 AL627248	AL627248 Danto rer
44	90.0	2.2	198520	2 AC129001	AC129001 Rattus no
45	90.0	2.2	205221	2 AC115723	AC115723 Mus muscu

ALIGNMENTS

RESULT 1

AX489468

LOCUS AX489468

DEFINITION Sequence 6768 from Patent WO02053728.

ACCESSION AX489468

VERSION AX489468.1 GI:22323480

KEYWORDS

SOURCE

ORGANISM

Candida albicans.

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE

1 Roemer, T., Jiang, B., Boone, C., Bussey, H., and Ohlsen, K.L.

Gene disruption methodologies for drug target discovery

Patent: WO 02053728-A 6768 11-JUL-2002;

AX489468 1794 bp DNA linear PAT 16-AUG-2002

Elitra Pharmaceuticals, Inc. (US)
 Location/Qualifiers
 Source 1. 1794
 /organism="Candida albicans"
 /db_xref="taxon:5476"
 BASE COUNT 706 A 286 C 322 G 480 T
 ORIGIN

Query Match 42.9%; Score 1767.2; DB 6; Length 1794;
 Best local similarity 99.4%; Pred. No. 7.9e-288;
 Matches 1784; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

134 ATGATTAATGCCAAGACACACATTAATCAACAAATATGAAAGAAATGAAACAAAGCA 193
 1 ATGATTAATGCCAAGACACACATTAATCAACAAATATGAAAGAAATGAAACAAAGCA 60
 194 TACAAATCAATATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 251
 61 TACAAATCAATATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 252 TCAGAAAGAAAACTGCCATGCTGAAGATGCCATGATGATGATGATGATGATGATGAT 311
 121 TCAGAAAGAAAACTGCCATGCTGAAGATGCCATGATGATGATGATGATGATGATGAT 180
 312 GCAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
 181 GCAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 372 TCAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
 241 TCAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 432 CTTCAGAAAGGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 491
 301 CTTCAGAAAGGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 492 GAATCTCAATTTGATGCTTTGTTTGAAACTCTAATGATTTGATGATGATGATGATGAT 551
 361 GAATCTCAATTTGATGCTTTGTTTGAAACTCTAATGATTTGATGATGATGATGATGAT 420
 552 CATGATCAATTTGATGCTTTGTTTGAAACTCTAATGATTTGATGATGATGATGATGAT 611
 421 CATGATCAATTTGATGCTTTGTTTGAAACTCTAATGATTTGATGATGATGATGATGAT 480
 612 AACGACCAAGATGCTTTGTTTGAAACTCTAATGATTTGATGATGATGATGATGATGAT 671
 481 AACGACCAAGATGCTTTGTTTGAAACTCTAATGATTTGATGATGATGATGATGATGAT 540
 672 AATGATGCTTTGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 731
 541 AATGATGCTTTGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 732 ATCAACCAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
 601 ATCAACCAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 792 GAAGATCAAAAAAGGAAAGAAATTAAGCAATTTGAAAGGCAAAAAATCTTCAAAAAGAT 851
 661 GAAGATCAAAAAAGGAAAGAAATTAAGCAATTTGAAAGGCAAAAAATCTTCAAAAAGAT 720
 852 GTTAAACATCTTGAAAGAAAGAAATTAAGCAATTTGAAAGGCAAAAAATCTTCAAAAAGAT 911
 721 GTTAAACATCTTGAAAGAAAGAAATTAAGCAATTTGAAAGGCAAAAAATCTTCAAAAAGAT 780
 912 TCAAGACATAGACATTTAATTTTCCAAATTTTCCAAAGTCTGATGATGATGATGATGAT 971
 781 TCAAGACATAGACATTTAATTTTCCAAATTTTCCAAAGTCTGATGATGATGATGATGAT 840
 972 TATCAAAAGACAGCAATGCTTGTCGAAAAATATATTAACCTTAACCTTAACCTTAACCTTA 1031
 841 TATCAAAAGACAGCAATGCTTGTCGAAAAATATATTAACCTTAACCTTAACCTTAACCTTA 900
 1032 AATCTTGAATTCGAGGTTGATCAAAAGAAAAATTTCAATTAAGAAAGTCTGATGATGAT 1091

901 AATCTTGAATTCGAGGTTGATCAAAAGAAAAATTTCAATTAAGAAAGTCTGATGATGAT 960
 1092 TCGTTGTCACAGCAAGCAAAAAATGTCACCAATATTAATTAATTAATTAATTAATTAAT 1151
 961 TCGTTGTCACAGCAAGCAAAAAATGTCACCAATATTAATTAATTAATTAATTAATTAAT 1020
 1152 AAAAATTTAGAAAGCAAAAGATCTTCTATTGATGATGATGATGATGATGATGATGATGAT 1211
 1021 AAAAATTTAGAAAGCAAAAGATCTTCTATTGATGATGATGATGATGATGATGATGATGAT 1080
 1212 AAACGATTTGACCTAATTTGACCAAGGTTTGTGATCAATTAATTCGAAAGATTTAGTCTAT 1271
 1081 AAACGATTTGACCTAATTTGACCAAGGTTTGTGATCAATTAATTCGAAAGATTTAGTCTAT 1140
 1272 ACCGCTGATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1331
 1141 ACCGCTGATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 1332 AAGCAATCAATGATGCTTTTGTCTCAACAATCCCTTGGACAAATGATTAACAGAAATAGCA 1391
 1201 AAGCAATCAATGATGCTTTTGTCTCAACAATCCCTTGGACAAATGATTAACAGAAATAGCA 1260
 1392 AAAATCGAAGATGACAAATAGCTAACCAACAATTAACCAAAACCAATAGTAATGTCAGAA 1451
 1261 AAAATCGAAGATGACAAATAGCTAACCAACAATTAACCAAAACCAATAGTAATGTCAGAA 1320
 1452 GAGGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1511
 1321 GAGGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 1512 ATGATGATCTTAATCTTGTATTTGTTGCTAGTAAGAAAAAGGATGATGATGATGATGAT 1571
 1381 ATGATGATCTTAATCTTGTATTTGTTGCTAGTAAGAAAAAGGATGATGATGATGATGAT 1440
 1572 GTTCCAGTACAGATTAATTTTATTAAGTTTCAACATTTCTAACGATCAAGAGAT 1631
 1441 GTTCCAGTACAGATTAATTTTATTAAGTTTCAACATTTCTAACGATCAAGAGAT 1500
 1632 GAATTAATTAAGAAAGAAATTAACCAACCAAAAGATCTCAATTTGATGATGATGATGAT 1691
 1501 GAATTAATTAAGAAAGAAATTAACCAACCAAAAGATCTCAATTTGATGATGATGATGAT 1560
 1692 GAACATTCAGTTCCCGCATGCTGATTAACAGACCTTATTAATTAAGTCAAACTTAGCACA 1751
 1561 GAACATTCAGTTCCCGCATGCTGATTAACAGACCTTATTAATTAAGTCAAACTTAGCACA 1620
 1752 GATGAACAGATCATTTCCATGACCAAGTGTTAATGTCAGACCTGATGATGATGATGATGAT 1811
 1621 GATGAACAGATCATTTCCATGACCAAGTGTTAATGTCAGACCTGATGATGATGATGATGAT 1680
 1812 TTTTCTTAATTAAGGATGCTGGAAGCGGAAAAAGCAAGGGGAAATCTTTGCAACAGAT 1871
 1681 TTTTCTTAATTAAGGATGCTGGAAGCGGAAAAAGCAAGGGGAAATCTTTGCAACAGAT 1740
 1872 TTTTCCAAAGCTAGGACCTTGACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1925
 1741 TCCTCCAAATTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1794

RESULT 2
 SCYR274C/c 3950 bp DNA 1linear PLN 11-AUG-1997
 LOCUS SCYR274C
 DEFINITION S.cerevisiae chromosome VII reading frame ORF YCR274C.
 ACCESSION Z73059.1
 VERSION Z73059.1 GI:1323498
 KEYWORDS
 SOURCE
 ORGANISM
 Saccharomyces cerevisiae.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetalia; Saccharomycetaceae; Saccharomycetes.
 REFERENCE
 1 (bases 1 to 3950)
 Panzeri, L., Agostoni Carbone, M.L., Melchiorretto, P., Plevani, P.,

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 09:41:49 ; Search time 5373 Seconds
(without alignments)
12406.593 Million cell updates/sec

Title: US-09-601-965C-1
Perfect score: 4116
Sequence: 1 aaaaaaaaaaagctccccc.....tattatgcattctcattgtc 4116

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estrov:*
- 6: em_estrpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_ges:*
- 18: em_ges_hum:*
- 19: em_ges_inv:*
- 20: em_ges_pln:*
- 21: em_ges_vrt:*
- 22: em_ges_fun:*
- 23: em_ges_mam:*
- 24: em_ges_mus:*
- 25: em_ges_other:*
- 26: em_ges_pro:*
- 27: em_ges_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227.8	5.5	907	17	CNS06URB AL416205 T7 end of
2	219.8	5.3	940	17	CNS075F8 AL430026 clone BAO
3	212	5.2	753	17	CNS06VZL AL417799 T3 end of
4	199	4.8	984	17	CNS06GB9 AL397483 T7 end of
5	116.8	2.8	891	17	AZ683582 ENTJK47TR
6	113.8	2.8	518	17	AZ928433 479.d1f09

7	108.2	2.6	912	17	AZ551092	AZ551092	ENTFJ22TF
8	107.2	2.6	877	17	AZ531291	AZ531291	ENTBQ34TR
9	105.6	2.6	816	17	AZ535744	AZ535744	ENTCQ25TR
10	104	2.5	997	17	CNS005TE	AL060767	DROSOPH11
11	101.2	2.5	908	17	AZ548467	AZ548467	ENTEK30TR
12	99.8	2.4	908	17	AZ551618	AZ551618	ENTDV54TR
13	99.8	2.4	906	17	BH153606	BH153606	ENTTS83TR
14	98.6	2.4	900	17	AZ53980	AZ549980	ENTDD94TF
15	96.2	2.3	976	17	BH149983	BH149983	ENTOD93TF
16	95.4	2.3	580	13	BJ433319	BJ433319	BJ433319
17	94.8	2.3	574	13	BJ433734	BH160272	ENTOV49TR
18	94.8	2.3	931	17	BH160272	AZ546009	ENTFW53TF
19	94.6	2.3	849	17	AZ546009	AZ676218	ENTK36TR
20	94.4	2.3	823	17	AZ676218	BQ731479	AGENCOURT
21	94.2	2.3	843	17	BH139532	BH139532	ENTNG88TF
22	92.8	2.3	843	13	BJ428840	BH146886	ENTPK48TF
23	92.4	2.2	890	17	BH146886	AZ550256	ENTEV58TR
24	92	2.2	905	17	AZ550256	BM415088	OP20159 M
25	91.4	2.2	483	13	BM415088	AZ528485	ENTCM64TF
26	91	2.2	934	17	AZ528485	AL263440	Tetraodon
27	90.8	2.2	795	17	CNS03W9J	AL226115	Tetraodon
28	90.8	2.2	838	17	CNS03W9J	AZ529191	ENTBV68TR
29	90.2	2.2	1135	17	CNS033CQ	BJ408820	BJ408820
30	90	2.2	880	17	AZ529191	BJ443268	BJ443268
31	89.6	2.2	589	13	BJ443268	AO051650	nbxb0002C
32	88.8	2.2	630	13	BJ443268	AZ441705	1M0234C01
33	88.4	2.1	705	17	AQ051650	BM415686	OP20766 M
34	88.4	2.1	739	17	AZ441705	BJ390371	BJ390371
35	88.4	2.1	966	13	BM415686	AU262882	AU262882
36	88.2	2.1	709	13	BJ390371	AZ279446	RPCI-23-1
37	88	2.1	461	9	AU262882	AZ015537	RPCI-23-3
38	87.2	2.1	527	17	AZ279446	AG044360	Pan trogl
39	87	2.1	592	17	AZ015537	AL250012	Tetraodon
40	86.6	2.1	694	17	AG044360	BM415494	OP20570 M
41	86.6	2.1	1036	17	CNS03LMT	B12981	T24D11-Sp6
42	86.6	2.1	948	13	BM415494	AZ461180	1M0266M19
43	86.4	2.1	1223	17	B12981		
44	86	2.1	495	17	AZ461180		
45	85.8	2.1					

ALIGNMENTS

RESULT 1
LOCUS CNS06URB 907 bp DNA linear GSS 06-JUL-2001
DEFINITION T7 end of clone AX0AA012E10 of library AX0AA from strain CBS 7064
ACCESSION AL416205
VERSION AL416205.1 GI:12196372
KEYWORDS GSS.
SOURCE Pichia farinosa.
ORGANISM Pichia farinosa.
REFERENCE Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Boletín-Pukhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,U., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Maupertuy,A., Neveuglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,M., Tekala,F., Toifano-Nicohe,C., Wesolowski-Louvel,M., Wincker,P. and Weissbach,J.
AUTHORS Yeast species for molecular evolution studies
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 1152876
REFERENCE 2 (bases 1 to 907)
AUTHORS de Montigny,U., Spehner,C., Souciet,J., Tekala,F., Dujon,B., Wincker,P., Artiguenave,F. and Potier,S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 09:51:39 ; Search time 134 Seconds
(without alignments)
9420.012 Million cell updates/sec

Title: US-09-601-965C-1
Perfect score: 4116
Sequence: 1 aaaaaaaaaaacgtccccc.....tattatgcatcttattgt 4116

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/6C_COMB.seq:*
6: /cgn2_6/prodata/1/ina/6D_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.2	2.8	7218	1 US-08-232-463-14	Sequence 14, Appl
2	75.2	1.8	696	4 US-09-461-697-193	Sequence 193, App
3	75.2	1.8	699	4 US-09-461-697-191	Sequence 191, App
4	75.2	1.8	717	4 US-09-461-697-189	Sequence 189, App
5	75.2	1.8	774	4 US-09-461-697-187	Sequence 187, App
6	75.2	1.8	819	4 US-09-461-697-185	Sequence 185, App
7	75.2	1.8	1669	4 US-09-461-697-184	Sequence 184, App
8	68.4	1.7	72604	4 US-09-268-992-7	Sequence 7, Appli
9	68.4	1.7	72604	4 US-09-657-474-7	Sequence 7, Appli
10	67.8	1.6	2223	4 US-08-257-073-4	Sequence 4, Appli
11	64.6	1.6	44453	4 US-09-146-053-5	Sequence 5, Appli
12	64.6	1.6	45546	4 US-09-146-053-6	Sequence 6, Appli
13	63.8	1.6	5394	3 US-08-688-376-1	Sequence 1, Appli
14	63.2	1.5	1559	4 US-09-019-095A-7	Sequence 7, Appli
15	62.6	1.5	240	4 US-08-628-417-6	Sequence 6, Appli
16	62.6	1.5	152311	3 US-09-128-155-16	Sequence 16, Appli
17	62.6	1.5	176313	3 US-09-128-155-17	Sequence 17, Appli
18	61.6	1.5	3211	2 US-08-574-859A-8	Sequence 8, Appli
19	61.6	1.5	3211	4 US-09-357-014-8	Sequence 8, Appli
20	61.6	1.5	3901	2 US-08-574-959A-6	Sequence 6, Appli
21	61.6	1.5	3901	4 US-09-357-014-6	Sequence 6, Appli
22	61.4	1.5	289	4 US-09-007-005-17	Sequence 17, Appli
23	61.4	1.5	289	4 US-09-244-796-17	Sequence 17, Appli
24	60.6	1.5	376	2 US-08-623-906A-18	Sequence 18, Appli
25	60.4	1.5	291	1 US-07-922-723A-7	Sequence 7, Appli
26	60.4	1.5	291	1 US-07-799-828C-7	Sequence 7, Appli
27	60.4	1.5	291	1 US-08-074-275-7	Sequence 7, Appli

28	60.4	1.5	291	1 US-08-480-366-7	Sequence 7, Appli
29	60.4	1.5	291	2 US-07-952-277A-7	Sequence 7, Appli
30	60.4	1.5	494	2 US-08-332-766A-22	Sequence 22, Appli
31	60.2	1.5	2169	4 US-09-434-408-3	Sequence 3, Appli
32	60.2	1.5	5935	4 US-09-178-973B-17	Sequence 17, Appli
33	60.2	1.5	5935	4 US-09-419-568F-29	Sequence 29, Appli
34	60.2	1.5	5935	4 US-09-354-243B-29	Sequence 29, Appli
35	59.8	1.5	2791	4 US-09-570-367C-1	Sequence 1, Appli
36	59.4	1.4	2447	2 US-09-014-969-14	Sequence 14, Appli
37	59	1.4	1236	4 US-08-741-134-5	Sequence 5, Appli
38	59	1.4	5361	4 US-08-973-462-2	Sequence 2, Appli
39	59	1.4	6152	4 US-08-973-462-1	Sequence 1, Appli
40	58.4	1.4	333	4 US-09-018-584A-27	Sequence 27, Appli
41	57.8	1.4	361	4 US-09-018-584A-9	Sequence 9, Appli
42	57.8	1.4	1298	3 US-08-948-705-3	Sequence 3, Appli
43	57.8	1.4	1447	4 US-09-443-041A-27	Sequence 27, Appli
44	57.6	1.4	43795	3 US-08-742-185-101	Sequence 101, App
45	57.4	1.4	282	4 US-09-461-697-205	Sequence 205, App

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29, 768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZcpT-F1s
US-08-232-463-14
Query Match 2.8%, Score 114.2, DB 1, Length 7218;

Best Local Similarity 11.6%; Pred. No. 2.2e-15; Matches 59; Conservative 271; Mismatches 179; Indels 0; Gaps 0;

Oy 2332 AATGTAAGTTTCAGAGTGAATGATAGGAGTGGACGATTTCTGTGAAGATG 2391
Db 1576 AATATCTAATTAATGAAGTGTGATGCTAGCAGTACGCTTCAAGCATCTA 1517
Oy 2392 TTTCACAACTTTCTGAGCATACGATATGCAAAATCGAAGATGAAGAATTGA 2451
Db 1516 TTTCAGTTTCAAAAAACGGCATGTACCTGTAATTAATCTATGCAAGTAGTA 1457
Oy 2452 TCGAATACCAAGACAGTGAAGGATGATGAAAGATTAAGAGATTAAGAGATG 2511
Db 1456 AAGAGATGAAGAAATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1397
Oy 2512 TTAATTCAGAGAAAGAGAAATTTGAACTATGATTAACACAGAAATTTCTATGATG 2571
Db 1396 RRR 1337
Oy 2572 ACACAAATGCAATTTGACAGCAGATAGATGATATATGTTTATTCGATGACAGA 2631
Db 1336 RRR 1277
Oy 2632 GCAGACAGATCATCTAGAATGTAAGAAAGTATTAAGAGAGATTCATTCGCGATG 2691
Db 1276 RRR 1217
Oy 2692 ATGCAAGAAATGAGATGATATTAATTAAGATTAAGAGAGATGAGAGAGAGAG 2751
Db 1216 RRR 1157
Oy 2752 AACAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAGAGAGAGAGAGAG 2811
Db 1156 RRR 1097
Oy 2812 ACAAAACCGAAGAGAGAAATCAAGAGAA 2840
Db 1096 RRR 1068

RESULT 2
US-09-461-697-193
Sequence 193; Application US/09461697
Patient No. 6277974
GENERAL INFORMATION:
APPLICANT: COGENT NEUROSCIENCE, Inc.
APPLICANT: Lo, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth
APPLICANT: Portbury, Stuart D.
APPLICANT: Putnam, Kasuri
APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461.697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 193
LENGTH: 696
TYPE: DNA
ORGANISM: Homo sapiens
US-09-461-697-193

Query Match 1.8%; Score 75.2; DB 4; Length 696;
Best Local Similarity 49.1%; Pred. No. 2.3e-07;
Matches 231; Conservative 0; Mismatches 233; Indels 6; Gaps 1;

Oy 2442 AAGAAATTTATGAAATACCAAGACAGGATGAGACCAAGGATTTGGAAGTAAAGAGA 2501
Db 130 AAGAAATTTATGAAATACCAAGACAGGATGAGACCAAGGATTTGGAAGTAAAGAGA 189

Oy 2502 TTAACGATGTAATTTCCAGAGAGAG-----AGAAATTCGAATCTAGATTACACAGAA 2555
Db 190 AAGAAAGAAAG 249
Oy 2556 GATTCCTGATGATGACACATGCAATTTGACAGCAAGTCTAGATGATTAATATG 2615
Db 250 GAAGATCAAAAGCAAG 309
Oy 2616 TTATTCGATGAG 2675
Db 310 GAATGAG 369
Oy 2676 GATTCATTCGAGATGATGACAGAAATGAGATGATTAATTAAGATTAAGAGAGAG 2735
Db 370 GATGAG 429
Oy 2736 GTTGAAG 2795
Db 430 GATGAG 489
Oy 2796 AAGACAAAG 2855
Db 490 AAG 549
Oy 2856 ACTGAATTTGATGTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2905
Db 550 AG 599

RESULT 3
US-09-461-697-191
Sequence 191; Application US/09461697
Patient No. 6277974
GENERAL INFORMATION:
APPLICANT: COGENT NEUROSCIENCE, Inc.
APPLICANT: Lo, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth
APPLICANT: Portbury, Stuart D.
APPLICANT: Putnam, Kasuri
APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461.697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 191
LENGTH: 699
TYPE: DNA
ORGANISM: Homo sapiens
US-09-461-697-191

Query Match 1.8%; Score 75.2; DB 4; Length 699;
Best Local Similarity 49.1%; Pred. No. 2.4e-07;
Matches 231; Conservative 0; Mismatches 233; Indels 6; Gaps 1;

Oy 2442 AAGAAATTTATGAAATACCAAGACAGGATGAGACCAAGGATTTGGAAGTAAAGAGA 2501
Db 133 AAGAAATTTATGAAATACCAAGACAGGATGAGACCAAGGATTTGGAAGTAAAGAGA 192
Oy 2502 TTAACGATGTAATTTCCAGAGAGAG-----AGAAATTCGAATCTAGATTACACAGAA 2555
Db 193 AAGAAAGAAAG 252
Oy 2556 GATTCCTGATGATGACACATGCAATTTGACAGCAAGTCTAGATGATTAATATG 2615
Db 253 GAAGATCAAAAGCAAG 312
Oy 2616 TTATTCGATGAG 2675

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 20:53:18 ; Search time 46 seconds
(without alignments)
3360.233 Million cell updates/sec

Title: US-09-601-965c-2

Perfect score: 6072

Sequence: 1 MEDLRDEALINGLNNNNNN.....CPLYKKWGLGIDDSAAVVG 1160

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6072	100.0	1160	20	AAV30128
2	1578	26.0	1066	19	AAW33632
3	1569	25.8	1066	17	AAW91300
4	765.5	12.6	1872	17	AAW06078
5	765.5	12.6	1872	18	AAW25030
6	765.5	12.6	1886	22	ABG06009
7	761.5	12.5	1575	23	AAW84357
8	760	12.5	1872	15	AAW56493
9	756	12.5	1893	15	AAW56491
10	756	12.5	1893	17	AAW06082

11	756	12.5	1893	18	AAW25020
12	756	12.5	1924	22	ABG06008
13	731.5	12.0	2065	22	ABW66356
14	693	11.4	1234	20	AAV30129
15	297	4.9	740	16	AAW68838
16	296	4.9	740	13	AAW27530
17	293	4.7	1193	21	AAW18306
18	286.5	4.8	1192	21	AAW18305
19	285.5	4.7	2485	21	AAW18172
20	283	4.7	1979	21	AAW18171
21	279	4.6	5024	22	AAW82935
22	278.5	4.6	10182	22	ABW38314
23	275.5	4.5	1812	22	ABW58022
24	269.5	4.4	1166	21	AAW18268
25	269	4.4	2206	21	AAW18254
26	263.5	4.3	1516	21	AAW18195
27	261.5	4.3	1780	22	AAW38681
28	259.5	4.3	1788	22	AAW40467
29	256.5	4.2	665	21	AAW18278
30	255.5	4.2	980	21	AAW18294
31	254	4.2	2285	20	AAW98149
32	254	4.2	6815	22	ABW66811
33	248	4.1	2688	22	AAW40883
34	246.5	4.1	4134	20	AAV31946
35	245	4.0	2663	22	AAW39097
36	244	4.0	1247	21	AAW18215
37	239.5	3.9	1052	22	AAW60276
38	237	3.9	2633	22	ABW06505
39	234	3.9	3696	23	ABP40235
40	233.5	3.8	2295	17	AAW18180
41	232.5	3.8	1089	17	AAW01896
42	232.5	3.8	1089	20	AAW05835
43	232.5	3.8	1089	21	AAW98055
44	231.5	3.8	2573	21	AAW18234
45	231	3.8	1351	21	AAW18290

ALIGNMENTS

```
RESULT 1
ID AAV30128 standard; Protein; 1160 AA.
XX
AC AAV30128;
XX
DT 14-OCT-1999 (first entry)
XX
DE A TAF-145 protein of Candida albicans.
XX
KW TAF-145 protein; protein-association factor; TAF-145; gene transcription;
XX fungal gene transcription; fungal infection.
XX
OS Candida albicans.
XX
PN WO940199-A2.
XX
PD 12-AUG-1999.
XX
PE 08-FEB-1999; 99WO-US02940.
XX
PR 09-FEB-1998; 98US-0024100.
XX
PA (SCRT-) SCRTPTGEN PHARM INC.
XX
PI Long F, Thompson CM, Wobbe RC;
XX WPI: 1999-494297/41.
XX DR N-PSDB; AA86650.
XX
PT Novel fungal protein critical for expression of fungal genes
XX provides a new target for antifungal drugs
```

PS Claim 7; Flg 3A-G; 115pp; English.

The pre-mRNA sequence represent a *Candida albicans* TAF-145 coding protein-associated factor (TAF) polypeptide, designated TAF-145. This protein is critical for the activated transcription of particular genes. TAF-145 polynucleotides may be used to identify other factors that interact with TAF-145. TAF-145 antibodies may be used to identify and quantitate TAF components, and to block the transcriptional function of TAF-145. Antibodies may also be used to identify, isolate and purify TAFs from various sources. The products may be used to identify agents useful in modulating fungal gene transcription. The compositions and agents are used for therapeutic purposes such as the treatment of fungal infections in mammals.

Sequence 1160 AA;

Query Match Similarity	100.0%	Pred. No. 0;	DB 20;	Length 1160;
Best Local Similarity	100.0%	Mismatches 0;	Indels 0;	Gaps 0
Matches 1160; Conservative				

Qy	1	MEIPREDEALINGLANSNNNNHDKRIDDDDBEFNLLEGOBELTINDEEMAAQAAAEOPFA	60
Dp	1	MEDLPDEALINGLANSNNNNHDKRIDDDDBEFNLLEGOBELTINDEEMAAQAAAEOPFA	60
Qy	61	LFQNSNDFDSNISHHDMGDSK:ITDDNHSSVNDHDLGFNNLGNGNHLLDDDNGND	120
Dp	61	LFQNSNDFDSNISHHDMGDSK:ITDDNHSSVNDHDLGFNNLGNGNHLLDDDNGND	120
Qy	121	LGELFFDDQEDSNVINTKRAKLDHDSNNDKTAQEDOKEREKNRQLKROKLOKIVKHEK	180
Dp	121	LGELFFDDQEDSNVINTKRAKLDHDSNNDKTAQEDOKEREKNRQLKROKLOKIVKHEK	180
Qy	181	EOIKRNLKXYFFPYRSRRRPENFIHFSPSPQOYRYORPALSKNLKPLPTVYNLEIEV	240
Dp	181	EOIKRNLKXYFFPYRSRRRPENFIHFSPSPQOYRYORPALSKNLKPLPTVYNLEIEV	240
Qy	241	DOKRTFKLRASDTASLSHEKKNV:NTIQODDLDFKLNESKRSSIDSEIKEDIVYKRDWTN	3000
Dp	241	DOKRTFKLRASDTASLSHEKKNV:NTIQODDLDFKLNESKRSSIDSEIKEDIVYKRDWTN	3000
Qy	301	CDKRDHYSKDLVSTTQWDDDATINACDNEYSIYKPINELLNNLPDNSQONOKIENDN	3600
Dp	301	CDKRDHYSKDLVSTTQWDDDATINACDNEYSIYKPINELLNNLPDNSQONOKIENDN	3600
Qy	361	TTNNYNNONNSNVODEEEDDDIFN:QITMLKRLKLDMPNPLLFPVSKKVADTKSVSESTK	4200
Dp	361	TTNNYNNONNSNVODEEEDDDIFN:QITMLKRLKLDMPNPLLFPVSKKVADTKSVSESTK	4200
Qy	421	LLELKENISNDQEYELLKKNYNTKORSQLSNLNIHSPVALLRLOTPTYKYVKLSTDETRSF	4800
Dp	421	LLELKENISNDQEYELLKKNYNTKORSQLSNLNIHSPVALLRLOTPTYKYVKLSTDETRSF	4800
Qy	481	HRPENVNRPGLVSFSKRLTKRKDKDGSKLQOOFKSTSDJLVADTGNITALEYSEQYP	5400
Dp	481	HRPENVNRPGLVSFSKRLTKRKDKDGSKLQOOFKSTSDJLVADTGNITALEYSEQYP	5400
Qy	541	ILSNFGMSGSKLINYYRKREPRMDT:SRPKAOIGETHIILVEBRSPPWNGGEVAPGDEVTVLY	6000
Dp	541	ILSNFGMSGSKLINYYRKREPRMDT:SRPKAOIGETHIILVEBRSPPWNGGEVAPGDEVTVLY	6000
Qy	601	NNMYRAPLFEKHDNKPOTPEFLVKS:GASASHOKFYLRGINFNPAVGNTPPVVAPAPHSKVT	6600
Dp	601	NNMYRAPLFEKHDNKPOTPEFLVKS:GASASHOKFYLRGINFNPAVGNTPPVVAPAPHSKVT	6600
Qy	661	NISKNRKLKMYFFRWYNSLGVRIYKVQYSKHFPHSDMONRQLKEPMELQROGEDGYN	7200
Dp	661	NISKNRKLKMYFFRWYNSLGVRIYKVQYSKHFPHSDMONRQLKEPMELQROGEDGYN	7200
Qy	721	KVRLGNDVIPGEEELRTMTTPED:SLMDTMOFGQOVLDDNNVLFGEOSROBSSSRKRGDK	7800
Dp	721	KVRLGNDVIPGEEELRTMTTPED:SLMDTMOFGQOVLDDNNVLFGEOSROBSSSRKRGDK	7800
Qy	781	REDISADDAENGDDIINKDKEREKSEKQOEKEEERKGDKEKDCKDEKDKTEKEKSKSKE	8400

Dd	761	RDDSDIADDAENDDINKKOEKEVEKKEDEREBEKKDKDEKDKEKDTEKEKSSKE	840
Qy	841	QDTEIDVEBELAPWNISRNFVIANOKRTIMOLGEBDDPGIGIGFSMLRATONKPFPLF	900
Dd	841	QDTEIDVEBELAPWNISRNFVIANOKRTIMOLGEBDDPGIGIGFSMLRATONKPFPLF	900
Qy	901	TPEPENVPKSNAAHNOKLYEDEIKRIWYSORSSLVDHGEGTSTKLQOITYNEPPRADHEL	960
Dd	901	TPEPENVPKSNAAHNOKLYEDEIKRIWYSORSSLVDHGEGTSTKLQOITYNEPPRADHEL	960
Qy	961	YUKNKLEDOQOOVOOOQODSLDAQDOOOOOOONRYLIRTRRVDENGIVHRKVEFIHPD	1020
Dd	961	YUKNKLEDOQOOVOOOQODSLDAQDOOOOOOONRYLIRTRRVDENGIVHRKVEFIHPD	1020
Qy	1021	RLIRAVVRKKKOIEDELKNADVDELPTNDKELTKRRKALEEKLANTLEBRAKOSRAKK	1080
Dd	1021	RLIRAVVRKKKOIEDELKNADVDELPTNDKELTKRRKALEEKLANTLEBRAKOSRAKK	1080
Qy	1081	PKPDLIHAAAAAGATIINDANTVWLPGGSYVIYGKGIGKGSRTPRCCNCAGAYGIIRINAK	1140
Dd	1081	PKPDLIHAAAAAGATIINDANTVWLPGGSYVIYGKGIGKGSRTPRCCNCAGAYGIIRINAK	1140
Qy	1141	CPLYKKMVLGIDDDSAAYVG 1160	
Dd	1141	CPLYKKMVLGIDDDSAAYVG 1160	

RESULT 2

AWW33632 standard; Protein; 1066 AA

AAW33632;

DT 06-JUL-1998 (first entry)

DE Yeast transcriptional activator factor TAF-145 protein.

214 TAF-145; transcriptional activator factor; transactivating factor;
215 KW

transcription; viability; antifungal; fungicide; infection; KW

XX
XX
XX

cinetapy.

XX
US
Saccharomyces cerevisiae

PN W09/49828-AL.
XX

31-DEC-1997.
PD
XX

26-JUN-1997
PE
XX

PR 26-JUN-1996; 9605-0673234
XX

PA (UYMA-) UNIV MASSACHUSETTS.
XX

PI Green MR, Reese JC;
yy

DR WPI; 1998-077189/07.

PT Yeast nucleic acids

PT patients

PS Example 1; Page 60-62; 93pp; English.

This polypeptide comprises a 145 kDa yeast transactivating factor, TAF-145, that is associated from *Saccharomyces cerevisiae* cells by virtue of its affinity to fungal or human TAFI-box binding proteins using chromatographic procedures. 2 peptides (see AW33630-31) obtained from TAF-145 were used to design degenerate oligonucleotides (see AA04774-75). These were used to screen a yeast genomic library, and an open reading frame was identified

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:27:24 ; Search time 32 Seconds
(without alignments)
3484.873 Million cell updates/sec

Title: US-09-601-965c-2

Perfect score: 6072

Sequence: 1 MEDLPKDEKINGLNNSNNNN.....CPYKKKVLGIDDDSAVVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_73:*

1: pir1:*\n2: pir2:*\n3: pir3:*\n4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1578	26.0	1066	2	S50237 TATA box-binding p
2	1537	25.3	979	2	S62469 probable transcrip
3	768	12.6	1865	1	I48155 transcription init
4	756	12.5	1893	1	A40262 transcription init
5	727	12.0	2068	2	A47371 transcription init
6	607	10.0	1490	2	S32373 DNA-binding protei
7	508.5	8.4	1994	2	D86452 protein F6N18.13 l
8	325	5.4	3394	2	T148501 hypothetical prote
9	297.5	4.9	1738	2	T14867 interaptin - slime
10	294	4.8	3724	2	T18427 hypothetical prote
11	293.5	4.8	743	2	A29232 10k malaria anti
12	289	4.8	1193	2	G71605 hypothetical prote
13	286.5	4.7	1192	2	A71623 probable secreted
14	285.5	4.7	2339	2	A45597 DNA-directed RNA p
15	285.5	4.7	2437	2	T18482 hypothetical prote
16	285.5	4.7	2485	1	H71621 serine/threonine-s
17	283	4.7	1979	2	C71622 hypothetical prote
18	282	4.6	4550	2	T18440 hypothetical prote
19	275	4.5	1526	2	A45605 mature-parasite-in
20	275	4.5	1790	2	S67593 transport protein
21	269.5	4.4	1166	2	H71609 hypothetical prote
22	269	4.4	2139	2	T18296 myosin heavy chain
23	269	4.4	2206	2	G71611 hypothetical prote
24	265	4.4	1780	2	T17272 hypothetical prote
25	264.5	4.4	1650	2	T18444 hypothetical prote
26	263.5	4.3	1127	2	T28317 ORF MSV156 hypothe
27	263.5	4.3	1516	2	E71619 RAD2 endonuclease
28	262	4.3	1871	2	D96796 probable heat choc
29	261.5	4.3	1312	1	BMVYDL RAD50 protein - ye

30	259	4.3	2401	2	T28676 rhotery protein -
31	258.5	4.3	1422	2	T18404 chromatin remodell
32	258.5	4.3	2910	2	T28156 DNA-directed RNA p
33	256.5	4.2	665	2	B71609 hypothetical prote
34	256	4.2	1658	2	S55101 hypothetical prote
35	255.5	4.2	980	2	E71606 hypothetical prote
36	254	4.2	2285	2	T12796 probable transglyc
37	250	4.1	1313	2	A48467 myosin heavy chain
38	250	4.1	1805	1	A64224 hypothetical prote
39	248.5	4.1	2022	2	T43214 ovrl protein - nem
40	248	4.1	2526	2	T20532 hypothetical prote
41	248	4.1	2722	2	E28032 hypothetical prote
42	248	4.1	2738	2	E88320 protein F07A11.6 l
43	246.5	4.1	1939	2	T18372 repeat organellar
44	246	4.1	2663	1	S28261 kinesin heavy chn
45	244.5	4.0	1254	2	T18277

ALIGNMENTS

RESULT 1

S50237
TATA box-binding protein-associated factor chain TAF1145 - yeast (Saccharomyces cere
N:Alternate names: protein G9374; protein YGR274c; TAF1130 protein
C:Species: Saccharomyces cerevisiae
C:Date: 27-Jan-1995 #sequence-revision 10-Feb-1995 #text-change 21-Jul-2000
C:Accession: S50237; S62389; S62397; S62394; S62395; S64607
R:Reese, J.C.; Apone, L.; Walker, S.S.; Griffin, L.A.; Green, M.R.
submitted to the EMBL Data Library, September 1994
A:Description: Yeast TAF115 in a multisubunit complex required for activated transcri
A:Reference number: S50237
A:Accession: S50237
A:Molecule type: DNA
A:Residues: 1-1066 <RE>
A:Cross-references: EMBL:U14954; NID:9595837; PID:9595838
R:Reese, J.C.; Apone, L.; Walker, S.S.; Griffin, L.A.; Green, M.R.
Nature 371, 523-527, 1994
A:Title: Yeast TAF115 in a multisubunit complex required for activated transcriptio
A:Reference number: S62389; MUID:95021683; PMID:7935765
A:Accession: S62389
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1066 <RE>
A:Cross-references: EMBL:U14954; NID:9595837; PID:AA979178.1; PID:9595838
A:Accession: S62397
A:Molecule type: protein
A:Residues: 583-599; 651-671 <RE>
R:Poon, D.; Bai, Y.; Campbell, A.M.; Bjorklund, S.; Kim, Y.J.; Zhou, S.; Kornberg, R.
Proc. Natl. Acad. Sci. U.S.A. 92, 8224-8228, 1995
A:Title: Identification and characterization of a TFIID-like multiprotein complex fro
A:Reference number: S62390; MUID:95396770; PMID:7667272
A:Accession: S62394
A:Molecule type: DNA
A:Residues: 370-830 <PO>
A:Accession: S62395
A:Molecule type: protein
A:Residues: 368-384; 528-554; 752-783 <PO>
R:Panerai, L.; Agostoni Carbone, M.L.; Melchiorretto, P.; Plevani, P.; Martegani, E.;
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64591
A:Accession: S64607
A:Molecule type: DNA
A:Residues: 1-1066 <PAN>
A:Cross-references: EMBL:273059; NID:91323498; PID:e243586; PID:91323499; MIPS:YGR274
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:TAF145; TAF130
A:Cross-references: SGD:S0003506; MIPS:YGR274c
A:Map position: 7R

Query Match 26.0%; Score 1578; DB 2; Length 1066;
Best local Similarity 32.4%; Pred. No. 1.4e-61;

Matches	391;	Conservative	204;	Mismatches	353;	Indels	260;	Gaps	37
QY	38	GOPETLNDEEMAAQAAASQAFALDG-----	11:11:11	-----NSNDPDSN	11:11:11	-----SHNH	77		
Db	8	GKTNLANDE-----AYEAIFGEGSGLEISGYIGDDEANSKDYTEHLPAVDPEDE	11:11:11	-----	11:11:11	-----	62		
QY	78	MGGDSNGCIDDNHHSVNDHD-----GFPNLGNGNHLLDDNDGN	11:11:11	-----DLGL	11:11:11	-----DLGL	124		
Db	63	LADDDDLPEE-----SDANLHPAMTMCAYDDVNGAVLGIDSNSIMOLPREINGLSQ	11:11:11	-----	11:11:11	-----	119		
QY	125	F---DD-----OOEDSVYN-----TKKIKL	11:11:11	-----	11:11:11	-----DDSN	148		
Db	120	FILEDDGCTPATSNALFPMGDANETHLETGVLGGSCANETIGHSOLSIGVNGNDMSIN	11:11:11	-----	11:11:11	-----	179		
QY	149	DGKTAOEDQEKENKROKLOKIVKLEEQIKRNIKYFPPTYSRHPENFK--FPS	11:11:11	-----	11:11:11	-----	207		
Db	180	GGFIMEPMSDCKHAKRKLD-----LINH-EVYLK	11:11:11	-----YEPPEKKILKWNLIYR	11:11:11	-----	230		
QY	208	PSPOYYRYQRAIALSKNIKPLIPTVNLIEVDOKIETKLSADTASLSHEDKVNIT	11:11:11	-----	11:11:11	-----	267		
Db	231	RSVPNHMSE-----ISRYKCPMPPLNKKRYOODKRLNSRTISVAPYIOGK--NLL	11:11:11	-----	11:11:11	-----	284		
QY	268	QDDLDIFIKLESKRSSIDSEFIKEYDKWDMNCKEFDH-----YSKDLASTDMDDAI	11:11:11	-----	11:11:11	-----	323		
Db	285	QSN-----SASARGLIIVSIDELFPKIEOOK--RKIHDEKTISEDLILADMDQEKI	11:11:11	-----	11:11:11	-----	338		
QY	324	INAGNEYSIYKPINELLNPNLDSKONROKTEINDNTNANNONNSNVQDEED----	11:11:11	-----	11:11:11	-----	378		
Db	339	INOGSSTATL-----ADSSMPTNLKF-----SGGX-KLKSLEIDYAEEDQWDE	11:11:11	-----	11:11:11	-----	381		
QY	379	DIIFGQATNLDC-LKLDNMDPMLLEVPSP-----KVDAATKSVAPETDKLELKFNI	11:11:11	-----	11:11:11	-----	428		
Db	382	DMIIAPKLESNAHLENNNDKELLMLIETKTNLAQOQKODSSNILLPLNETHILOOKFNL	11:11:11	-----	11:11:11	-----	441		
QY	429	SNDQYELLRKRYNFKORSOLSNLNIHSVPAIRLOTPTYKVKLSIDFETRSHPRPV--N	11:11:11	-----	11:11:11	-----	486		
Db	442	SNDQYQILKTKHQKVKVSTISNLMHOSOPALINQSPRYKAVVRQYQHRHREFGSH	11:11:11	-----	11:11:11	-----	501		
QY	487	VRPGLVFSFSKILKRRKKDKSLQOAFSKSLDTLVADTGNIILVEYEOYPRILSNG	11:11:11	-----	11:11:11	-----	546		
Db	502	IRPGKGIYFSKLAKKRRKRDGKDYKESSTQDITIGTAVYILMEYSEQFPVALSKG	11:11:11	-----	11:11:11	-----	561		
QY	547	MGSKLINTYRKREPRDTSRPKAQIGETHILGVEDNSPFWNGEVAABGFVPLLYNNMVA	11:11:11	-----	11:11:11	-----	606		
Db	562	MANKLINTYRKREPRDTSRPKAQIGETHILGVEDNSPFWNGEVAABGFVPLLYNNMVA	11:11:11	-----	11:11:11	-----	621		
QY	607	PIFKHDNKPRTDFLAVKSOAGASHQKPYFLGILNFAVNGTFPV-EVPAHSHKVTNISKN	11:11:11	-----	11:11:11	-----	665		
Db	632	PVKFHDISGIDTLKLRKSSGFGISNFPYLNINHLITLVGGTFPVEELPGNSKRYVSMKAT	11:11:11	-----	11:11:11	-----	681		
QY	666	RLKAVYFVWMSLSLVPKISVDVSKNHPEDSHDMONRORLKEMEYOROGEDGYKVRGL	11:11:11	-----	11:11:11	-----	725		
Db	682	RLKMLIYRILNINHSKALSIDPILAHFPDODYGONRQKKEKMKYORORDPEKGLMKLS-	11:11:11	-----	11:11:11	-----	740		
QY	726	NDVIJPGEEIIRTMPEJDSLMDIQFOGOVYLDMMVYLFGEOSROESSRSRKGDREDSI	11:11:11	-----	11:11:11	-----	785		
Db	741	DEKLNDIAVYSLIRPEOISQVESMSQIQOFEDN-----	11:11:11	-----	11:11:11	-----	775		
QY	786	ADDAENCGDINDKKEEVEKEDEEREKEKDKKKDKKDKTEKEKSKKEDOTEI	11:11:11	-----	11:11:11	-----	845		
Db	776	--EAVNFP-----SKLS-----	11:11:11	-----	11:11:11	-----	786		
QY	846	DVEEELAAWNLSRNVIANOTKTMOLNGEGDPTGIGFSMLRATOKRKPRLPTPPE	11:11:11	-----	11:11:11	-----	905		
Db	787	LEEBLLRPMNITKTINISTOMRAMQIHOVGDPTCGCEGFSFLKTSKMGCGFKSGSPSN	11:11:11	-----	11:11:11	-----	845		
QY	906	NVPR-----SNAANQKLYBOEIKRIVYQSSSLVLDHGEESKILQOITYNEPPAD	11:11:11	-----	11:11:11	-----	957		
Db	846	NNSSKKGTNTHTSYVAQQAQKAYDEEIAKTYHTHS-----	11:11:11	-----	11:11:11	-----	882		
QY	958	HELYLKNKLEDOQVOQOQDPSLOADQOQOQOQOQOONRYLRTIRAYRDNGLIHKRVEFI	11:11:11	-----	11:11:11	-----	1017		
Db	883	--LISNFE-----EMTNPDEI NOTNKHVKTRDRDKLTKLVKRRKRDENGITIGQTFE	11:11:11	-----	11:11:11	-----	935		

[illegible]

RESULT 2

C: probable transcription initiation factor tflid subunit - fission yeast (Schizosaccharomyces pombe)
 C: Species: Schizosaccharomyces pombe
 C: Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jan-2000
 C: Accession: T38580; 562463
 R: Badoock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, October 1995
 A: Reference number: Z21745
 A: Accession: T38580
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-979 <BA2>
 A: Cross-references: EMBL:Z54354; NID: g1019398; PIDN: CAA91179.1; PID: g1019412; GSPDB: G1019412
 A: Experimental source: strain 972h-; cosmid c2G11
 C: Genetics:
 A: Gene: SPAC2G11.14
 A: Map position: 1L
 A: Introns: 21/3; 212/1; 261/3
 C: Keywords: transcription initiation

Query Match	Score	DB 2	Length
25.3%	1537	2	979

Matches 374; Conservative 196; Mismatches 349; Indels 262; Gaps 31;

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:01:08 ; Search time 21 Seconds
(without alignments)
2291.073 Million cell updates/sec

Title: US-09-601-965c-2

Sequence: 1 MEDLPDEAINGLNSNNNN.....CPLKKMVLGIDDSAAVVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1578	25.0	1066	T145_YEAST	P04677; T145_YEAST
2	1537	25.3	979	T111_SCHPO	Q09813 schizosacch
3	765.5	12.6	1872	T2D1_HUMAN	P21675 homo sapien
4	723	11.9	2068	T2D1_DROME	P51123 drosophila
5	293.5	4.8	743	ABRA_PLAFC	P22620 plasmodium
6	285.5	4.7	2339	RCPI_PLAFA	P27625 plasmodium
7	274.5	4.5	1790	USO1_YEAST	P23386 saccharomyc
8	261.5	4.3	1312	RA50_YEAST	P12753 saccharomyc
9	256	4.2	1658	YM67_YEAST	Q03661 saccharomyc
10	250	4.1	1805	HMM2_MYCCE	P47460 mycoplasma
11	248.5	4.1	2022	ANT1_ONCYO	P21249 onchocerca
12	246	4.1	2663	CENE_HUMAN	Q02224 homo sapien
13	242.5	4.0	1875	MLP1_YEAST	Q02455 saccharomyc
14	235	3.9	1332	SPY7_YEAST	P35177 saccharomyc
15	233.5	3.8	1938	MYSD_CAEEL	P02567 caenorhabd
16	233.5	3.8	2869	RBP1_PLAFA	Q00798 plasmodium
17	233	3.8	678	GARP_PLAFA	P13816 plasmodium
18	232.5	3.8	1089	NMD2_YEAST	P38798 saccharomyc
19	227.5	3.7	2245	MYSJ_DICDI	P54697 dicyostell
20	225	3.7	2230	GOG4_HUMAN	Q13439 homo sapien
21	224	3.7	1928	MYSL_YEAST	P08964 saccharomyc
22	223	3.7	2349	TPR_HUMAN	P12270 homo sapien
23	222	3.7	1233	YF16_YEAST	P43597 saccharomyc
24	222	3.7	1957	YD86_SCHPO	Q10411 schizosacch
25	220.5	3.6	1664	YIO9_YEAST	P53705 candida alb
26	220.5	3.6	1664	YIO9_YEAST	P40457 saccharomyc
27	219	3.6	1085	IFN1_YEAST	P39520 saccharomyc
28	218.5	3.6	1340	YNU1_YEAST	P53935 saccharomyc
29	217.5	3.6	899	YMY3_YEAST	Q04500 saccharomyc
30	217.5	3.6	3210	MYHB_CHICK	P10587 gallus gall
31	217.5	3.6	3210	CENE_HUMAN	P49454 homo sapien
32	216.5	3.6	1898	TRHY_HUMAN	Q07283 homo sapien
33	216.5	3.6	1938	MYA_AEOIR	P24733 aequipecten

34	216.5	3.6	1976	1	MYHA_BOVIN	Q27991 bos taurus
35	214	3.5	2871	1	DESP_HUMAN	P15924 homo sapien
36	211.5	3.5	1324	1	CUT3_SCHPO	P41004 schizosacch
37	211.5	3.5	1630	1	MSPI_PLAFA	P04932 plasmodium
38	211.5	3.5	1639	1	MSPI_PLAFA	P04933 plasmodium
39	211.5	3.5	1818	1	HMM2_MYCEN	P73471 mycoplasma
40	211.5	3.5	2116	1	MYSD_DICDI	P08799 dicyostell
41	211	3.5	1962	1	MYSA_DROME	P05661 drosophila
42	210.5	3.5	1972	1	MYHA_MOUSE	Q08638 mus musculu
43	209.5	3.5	1513	1	STU1_YEAST	P38198 saccharomyc
44	209.5	3.5	1976	1	MYHA_RAT	Q91150 rattus norv
45	207	3.4	1178	1	MNNA_YEAST	P36044 saccharomyc

ALIGNMENTS

RESULT 1
T145_YEAST
ID T145_YEAST STANDARD: PRT: 1066 AA.

AC P46677; 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 35, Last annotation update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Transcription initiation factor TFIID 145 kDa subunit (TBP-associated factor 145 kDa) (TAFII-145) (TAFII-130).
DE factor 145 kDa (TAFII-145) (TAFII-130).
GN TAFI145 OR YGR274C OR G9374.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RX [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 583-599 AND 651-671.
RC STRAIN=x57;
RC MEDLINE=95021683; PubMed=7935765;
RX Reese J.C., Apone L., Walker S.S., Griffin L.A., Green M.R.;
RT "Yeast TAFIIS in a multisubunit complex required for activated transcription.";
RT Nature 371:523-527(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RC MEDLINE=97279232; PubMed=9133740;
RX Ruzzi M., Marconi A., Sallola M., Fabiani L., Montebove F.,
RA Frontali L.;
RT "The sequence of a 8 kb segment on the right arm of yeast chromosome VII identifies four new open reading frames and the genes for yeast TAFI145.";
RT Yeast 13:365-368(1997).
RN [3]
RP SEQUENCE OF 368-384; 528-554 AND 752-783, AND CHARACTERIZATION.
RC STRAIN=YPH252;
RC MEDLINE=95399770; PubMed=7667272;
RX Poon D., Bai Y., Campbell A.M., Bjorklund S., Kim Y.-J., Zhou S.,
RA Kornberg R.D., Weill P.A.;
RT "Identification and characterization of a TFIID-like multiprotein complex from Saccharomyces cerevisiae.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:8224-8228(1995).
RT FUNCTION: TAFs ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID (TFIID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF RNA POLYMERASE TRANSCRIPTION.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs) WHOSE MW RANGE FROM 25-150 kDa.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: TO S. POMBE TAFII-11. SOME, TO HUMAN TAFII-250 (CCG1) AND TO DROSOPHILA TAFII-230.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:26:54 ; Search time 50 Seconds
(without alignments)
4780.293 Million cell updates/sec

Title: US-09-601-965C-2
Perfect score: 6072
Sequence: 1 MEDIPRDEAINGLNNNNNN.....CPLYKKVGLGIDDSAAVVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-virus:*
- 16: sp-bacteriophage:*
- 17: sp-archaeal:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	768	12.6	1865	11 060544	060544 mesocricetu
2	760.5	12.5	883	5 08SR44	08SR44 encephalito
3	731.5	12.0	2065	5 097068	097068 drosophila
4	607	10.0	1490	5 09TX96	09TX96 drosophila
5	561.5	9.2	1792	5 09XUL9	09XUL9 caenorhabdi
6	536.5	8.8	1810	10 09LJ62	09LJ62 arabidopsis
7	508.5	8.4	1994	10 09LPI9	09LPI9 arabidopsis
8	339.5	5.6	1946	5 09J291	09J291 plasmodium
9	328	5.4	3130	5 09BK46	09BK46 plasmodium
10	325	5.4	3394	5 077384	077384 plasmodium
11	308	5.1	2867	5 09N2M3	09N2M3 plasmodium
12	300	4.9	959	5 08TIP2	08TIP2 dictyosteli
13	298	4.9	440	11 035361	035361 mus musculu
14	287.5	4.9	1387	5 09G276	09G276 plasmodium
15	287.5	4.9	1738	5 076329	076329 dictyosteli
16	295.5	4.9	3254	5 09BK45	09BK45 plasmodium

17	294	4.8	3724	5 077320	077320 plasmodium
18	291.5	4.8	2081	10 09LH98	09LH98 arabidopsis
19	291	4.8	1461	5 08ST04	08ST04 dictyosteli
20	290.5	4.8	1791	5 08T207	08T207 dictyosteli
21	289	4.8	1193	5 096257	096257 plasmodium
22	286.5	4.7	1114	5 097242	097242 plasmodium
23	285.5	4.7	1192	5 096127	096127 plasmodium
24	285.5	4.7	2423	5 077393	077393 plasmodium
25	285.5	4.7	2485	5 096134	096134 plasmodium
26	285	4.7	1699	5 095PH3	095PH3 dictyosteli
27	283	4.7	1979	5 096133	096133 plasmodium
28	282	4.6	4550	5 077336	077336 plasmodium
29	280.5	4.6	739	5 09N1H1	09N1H1 plasmodium
30	278	4.6	1661	5 006166	006166 plasmodium
31	275.5	4.5	1812	5 09VZD9	09VZD9 drosophila
32	275	4.5	1510	5 025920	025920 plasmodium
33	275	4.5	1790	5 007380	007380 plasmodium
34	271.5	4.5	2771	5 026216	026216 plasmodium
35	270	4.4	3628	5 0968Y0	0968Y0 plasmodium
36	269.5	4.4	1166	5 096219	096219 entamoeba
37	269	4.4	2139	5 006205	006205 plasmodium
38	268.5	4.4	2206	5 096205	096205 plasmodium
39	268.5	4.4	16215	5 09NFS3	09NFS3 drosophila
40	265	4.4	1780	4 090FR5	090FR5 homo sapien
41	265	4.4	2038	5 0967Y0	0967Y0 dictyosteli
42	264.5	4.4	1650	5 077328	077328 plasmodium
43	263.5	4.3	1127	12 09YVT6	09YVT6 melanoplus
44	263.5	4.3	1516	5 096154	096154 plasmodium
45	263.5	4.3	1583	4 015045	015045 homo sapien

ALIGNMENTS

RESULT 1

ID	060544	PRELIMINARY:	PRT:	1865 AA.
AC	060544;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CCG1.			
OS	Mesocricetus auratus (Golden hamster).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;			
OC	Mesocricetus.			
OX	NCBI_TaxID=10036;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94215915; PubMed=8163200;			
RA	Hayashida T., Sekiguchi T., Noguchi E., Sunamoto H., Ohba T.,			
RA	Nishimoto T.;			
RT	"The CCG1/TAIR1250 gene is mutated in thermosensitive GI mutants of			
RT	the BHK21 cell line derived from goldenhamster.";			
RL	Gene 141:267-270(1994).			
DR	EMBL; D26114; BAA05110.1; -			
DR	InterPro; IPR001487; Bromodomain.			
DR	Pfam; PF00439; Bromodomain; 2.			
DR	PRINTS; PRO0503; Bromodomain.			
DR	SMART; SM00297; BROMO; 2.			
DR	PROSITE; PS00633; BROMODOMAIN_1; 2.			
DR	PROSITE; PS00633; BROMODOMAIN_2; 2.			
SO	SEQUENCE 1865 AA; A81614946C0C0F24 CRC64;			

Query Match 12.6% Score 768; DB 11; Length 1865;

Best Local Similarity 22.1%; Pred. No. 5.6e-28;
Matches 306; Conservative 226; Mismatches 445; Indels 408; Gaps 51;

QY	43	TNDEEAAQAAMAEQDFLFGNSND-----FDSNISHH-----	75
DB	18	TSDSESSGCGPPSLTGLFNGINGAGOLEGESVLDDECKKHLAGALGSLITELTA	77
QY	76	-DHMGDSNGIIDDN-----HHSSVND-----HDLFNNLGNHLLDDND	116

Db	78	NBIJLTDGALVNDGEWIRSRKEDAVDYSIDINEVAEDSRXYOOTMKSLOPLCHSABDDED	137
Oy	117	GLUJDLCELEP-----DOEEDSNVINTXKHKHKLDDSDNNDGKTAOEDOKENK	162
Db	138	DYDADCEDEDIDCKLMPRPBPYGVGWKKKEDODGLTGEXVDFSSSDSDSEMGPOEAOAES	197
Oy	163	KROIKRQKLOKIYKHLKEEDOKINIKITYPTTYSRHRNFHKFFSPS--POYRYQORA	219
Db	198	KOGKLTPLRPLAGIMQH-DARKILPSVTELEPPEFPRGKVLRLRLEGGKXVYWRARRK	256
Oy	220	ILSKNIKYLPIPTVY-----NLEIEVDOKIKF-----LRSa	251
Db	257	-RKKKHREPIOEJOIEECSVELEVNOKSLMWYDAPRPPPEOCLSDSEITMMARVESK	315
Oy	252	DYASLSHDEK-----NVTNITODDLDEFTKLESKRSSIDSEIKET	291
Db	316	FSQSGTDDKVDMDYTPRAEMRYGPARLWYDMLGVPEDSGDYGKMKRTEHAPAKCK	375
Oy	292	DYVK-RDMTNCDEKDHYSKD--LVLTSTDWDDAIINAGDNEYSTKRP-----INEL	340
Db	376	MWTKRKLEESGIDLLADENFLWLTQJLWEMDDIITWGDGEDYKHKGTQKPROASLAGIPLSS	435
Oy	341	LLNPNLDSKONRQKTEIENDNTYNNY-----QNSNVQOEEDDIFNCG-----INDKL	391
Db	436	MTRNMAAVNVOGFAATLDDDKPWSIFPIDNEDLYGWMENIITWDQAMRILEPYL	495
Oy	392	KLDNMDPMLIF-VPSAKVDATKSVYPSFDK-----LL-----E	423
Db	496	TLDPNDEDLILEIPDEKBEATSNPSKKNKKESSLAKSRILLGKTGVYKEEPOOMNSPE	555
Oy	424	LK--FNISNDEVELLRKNYNNKO--RSQSLNLEHNSVPLRLQTPRYKXKLSTDETRS	479
Db	556	VKDEMNLSNDEY-----YKOGGLKGTPEGNIIOHSIRVAVELRYOPRPYIMGBIKIRO	609
Oy	480	FHRVVF-----N/RGCTLVSEFKLLKLRKKKDKGKSLQOIFSKT--SDLTVA	524
Db	610	FHRPRLKXSGALSGOPRSHVQOLKHKIKKAKMARDEKQASGGGEFMRTPQDLTLG	669
Oy	525	DTGNIILYESQYPLLSNFGMSKSLINYRKERPNDTSRPAKOIGET--HILGEDR	581
Db	670	D-DGLILAESEENGRLMVOYGMATKTKNNYKRRPGKDPGAPDCKEYELVYCH-----T	722
Oy	582	SPWNNGEVAPRGDFVPLLNNVRAPIFKHDNKPFTDLVKVQSQAGSHQKFYLRGINFNE	641
Db	723	SPF--GSLHPRGOLQAFENNIRAPRIYTLHKMPEITDLIRTR-----QGYIRELVDF	775
Oy	642	AVGNTFPV-EVPAHSHKRYTNI SKNRILKAVYFRM--RSLGVP--RISYKDVSKHPRHSOM	698
Db	776	VWQOQCPLEFVPPGNSKRANTHIRDFLOVPIRLFWKSKDRPRIRNEDIKKAFPSISES	835
Oy	699	QNBORLKEFEXXROGEDOYKVRGINDVYIPDEELFRITGPEEDSLMDTMOQGOYLD	758
Db	836	SJRKRLKLADFEKRTMDSMNWLKS--DFRLPTEELRAVNSPECCAIYSMAAEGRK	894
Oy	759	DNWVLTEGOSROESSRSRKGDKREDSIADAENGDDIINDKEVEKEKEQREEEKGD	818
Db	895	D--AGYGEKS-----FFABEENEBEP-----	914
Oy	819	KEKDOKEKDKTEKESKSKSEJDEITDYEBELAPRNI SRNFVIANOTYTMLOLNGEDP	878
Db	915	-----QMKIDEVARTAPWNTTRAPLTAARKGCLLEVGVADP	951
Oy	879	TGIGIGFSLRATOKNPKRPLETPPENVPKS-----NAAHNIQ--LYEO	922
Db	952	TGCGGCFSTYVKLPNK--PTQOKDKKEBPQVKKVTGTADLRLRLSLKNAQOLRKGFPVEE	1010
Oy	923	EKRLWYOSRSLV-----HGEGTSKLOQIYNEKPRADH--ELY-----	961
Db	1011	EKKR1-----SRMEVIDVVRTMST1QARSCEGSPKSPAR--GSRFSVAHEHGRXYKEECORIFD	1066
Oy	962	LKNLEDOOV-----OQOQDDPDLQDDQOQOQOQOQNR	995

```

Db      1067 LONKVLSPFEVSLSTDPDSSAEDSPFEENGKNIENMLONKKTSSOLSREBEROEKELOR 1126
Oy      996 v-----LRTTRRVNDENGIVHKKVEFIHDPRLL 1023
Db      1127 MLLAGSASAGNNHDDDDPASTVSLNSATGCGCLTYRFRREEKEVEYCRCTVKKRPAYI 1186
Oy      1024 RAYVR-RKKQIIDECLKNADVDEILLPTNDKELNKTIRRALAEKLANLEKRAKOSAKKPP 1082
Db      1187 DAYVIRITTKDEEFERKRALFEDE--QHHEERKKRRRR-IOBLRRLKNOBKEKLCGP 1242
Oy      1083 KDLHIAAAGATTIDANTVMLPDGSYVIGGKIGKIGKRSR--TRCKNCGAYGIRITNA 1139
Db      1243 E-----KKPKMKKERPDCLKCGACGCAIGMRTNK 1272
Oy      1140 KCPLY 1144
Db      1273 FCPLY 1277

RESULT 2
O8SR44 ID O8SR44 PRELIMINARY; PRT; 883 AA.
AC O8SR44;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Transcription Initiation factor TFIID 11kDa subunit.
OS Euc01.0760.
OS Encephalitozoon cuniculi
OC Eukaryota
OC NCB1_TaxID=6035;
RN RP SEQUENCE FROM N.A.
RA Genoscope;
RA STRAIN=GB-M1;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RC MEDLINE=21576510; PubMed=11719806;
RA Kariina M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Preissler G., Barbe V., Peyretallade E., Brotier P., Wincker P.,
RA Dellac F., El Aloui H., Peyret P., Saurin W., Gouy M.,
RA Vissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RL Nature 414:450-453(2001).
RL EMBL: AL590449; CAD25795.1; -.
SQ SEQUENCE 883 AA; 10200 MW; 799568726B6F45CD CRC64;

Query Match 12.5%; Score 760.5; DB 5; Length 883;
Best Local Similarity 24.7%; Pred. No. 5,3e-28;
Matches 253; Conservative 149; Mismatches 304; Indels 319; Gaps 35.

Oy      150 GKTAADEQKREKNRQKLRQKQKIVKHEKQIRNRIKITYPTYSRHRPFFHKFFSPS 209
Db      10 GRAADDEKSYSEAGDGPGRSGMVAIRFDDLLVIEKPYEYKLRKKGRK----- 58
Oy      210 PQYRQKRPALASLNIRPLIPTKYNLEIVDQKIKF-----LRSADTA-SLSHEDKN 262
Db      59 -----VPNE-SMSYEVDQDKFFNSMTYARDSNTTLOALSGSTRS 97
Oy      263 VTNIQTDDLDPIKNE-----SKRSSIDSFTKEIDYVRKDWNTCDKFHYSKDLVLTSD 317
Db      98 -TGLQO-----YLRKQLEKEISKRRALDSDSLREVEEVK-----YFDDIV---E 137
Oy      318 WDDDAIINAGDNEYSIYKPINELLNPNLDNSKONROKRIENDNTNNYNQNSNVODEEE 377
Db      138 WE-----NNIYDFDAGGRMKMDVTTER-----VDSILEDM 171
Oy      378 DDDIFNGQINLDR---LKDMDNPNLLEFVSKKVDVATKTSVVP-----STPKLELK 425
Db      172 ERYVYVDESOMSKRSKFLTYLEEDPNLIF---EKIDRKQAKPKPKKQKQFVSNYKPKK 726

```

```

Db      1067 LONKVLSPFEVSLSTDPDSSAEDSPFEENGKNIENMLONKKTSSOLSREBEROEKELOR 1126
Oy      996 v-----LRTTRRVNDENGIVHKKVEFIHDPRLL 1023
Db      1127 MLLAGSASAGNNHDDDDTASTVSLNSATGCGCLTYRFRREEKEVEYCTVKKRPAYI 1186
Oy      1024 RAYVR-RKKQIIDECLKNADVDEILLPTNDKELNKTIRRALAEKLANLEKRAKOSAKKPP 1082
Db      1187 DAYVIRITTKDEEFERKRALFEDE--QHHEERKKRRRR-IOBLRLRLKNODEKELKCP 1242
Oy      1083 KDLHIAAAGATTIDANTVMLPDGSYVIGGKIGKGRS--TRCKNCGAYGIRITNA 1139
Db      1243 E-----KKPKMKKERPDCLKCGACGCAIGMRTNK 1272
Oy      1140 KCPLY 1144
Db      1273 FCPLY 1277

RESULT 2
O8SR44
ID      O8SR44      PRELIMINARY;      PRT;      883 AA.
AC      O8SR44;
DT      01-JUN-2002 (TEMBLrel. 21, Created)
DT      01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE      01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE      Transcription Initiation factor TFIID 11kDa subunit.
OS      Euculio_0760.
OC      Encephalitozoon cuniculi
Ox      Eukaryota
NCBI_TaxID=6035;

RN      RP      SEQUENCE FROM N.A.
RA      Genoscope;
RA      STRAIN=GB-M1;
RL      Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
RP      (2)
RP      SEQUENCE FROM N.A.
RC      STRAIN=GB-M1;
RC      MEDLINE=21576510; PubMed=11719806;
RA      Katalina M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA      Preissler G., Barbe V., Peyretallade E., Brotier P., Wincker P.,
RA      Dellac F., El Aloui H., Peyret P., Saurin W., Gouy M.,
RA      Vissenbach J., Vivares C.P.;
RT      "Genome sequence and gene compaction of the eukaryote parasite
RT      Encephalitozoon cuniculi."
RL      Nature 414:450-453(2001).
EMBL: AL590449; CAD25795.1; -.
SQ      SEQUENCE 883 AA; 10200 MW; 799568726B6F45CD CRC64;

Query Match      12.5%; Score 760.5; DB 5; Length 883;
Best Local Similarity 24.7%; Pred. No. 5,3e-28;
Matches 253; Conservative 149; Mismatches 304; Indels 319; Gaps 35.

Oy      150 GKTAADEDEKREKNRQKLRQKQKIVKHEKQIRNRIKIYPTYSRHRPFNFHKFFSPS 209
Db      10 GRAADDEKSYSEAGDGPGRSGMVAIRFDLLVIEKPYEYKLRKKGRK-----58
Oy      210 PQYRQKRPALASLNIRPLIPTKYNLEIVDQKIFK-----LRSADTA-SLSHEDKN 262
Db      59 -----VPE-SMSVEYDDQKRFNSMTYARDSNTTLOALSGSTRS 97
Oy      263 VTNIQTDDLDPTKLE-----SKRSSIDSFTKEIDYVRKDWNTCDKFHYSKDLVLTSD 317
Db      98 -TGLQO-----YLRKQLEKEISKRRALDSDSLREVEEK-----YFDDV-----E 137
Oy      318 WDDDAIINAGDNEYSIYKPINELLNPNLDNSKONROKRIENDNTNNYNQNSNVODEEE 377
Db      138 WE-----NNIYDFDAGGRMKMDVTTER-----VDSILEDM 171
Oy      378 DDDIFNGQINLDR-----LKDMDNPNLLEFVSKKVDVATKSVP-----STPKLELK 425
Db      172 ERYVYVDESOMSKRSKFLTYLEEDPNLIF-----EKIDRKQAKPKPKOKFVSNYKPKK 726

```

Wed Mar 12 10:13:17 2003

us-09-601-965c-2.ra1

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:26:03 : Search time 22 Seconds
(without alignments)
1551.390 Million cell updates/sec

Title: US-09-601-965C-2
Perfect score: 6072
Sequence: 1 MEDLPDEALINGLNNNNNN.....CPLYKKVGLGIDDSAAVVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1569	25.8	1066	2 US-08-308-818-1	Sequence 1, Appl
2	765.5	12.6	1872	1 US-08-188-582-14	Sequence 14, Appl
3	765.5	12.6	1872	1 US-08-646-715-14	Sequence 14, Appl
4	756	12.5	1893	1 US-08-188-582-11	Sequence 11, Appl
5	756	12.5	1893	1 US-08-646-715-11	Sequence 11, Appl
6	722	11.9	1377	2 US-08-308-818-4	Sequence 4, Appl
7	297	4.9	740	1 US-08-237-073-5	Sequence 5, Appl
8	278.5	4.6	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
9	254	4.2	2285	4 US-09-308-375-2	Sequence 2, Appl
10	234	3.9	3696	4 US-09-134-001C-5080	Sequence 5080, Ap
11	232.5	3.8	1089	1 US-08-375-300-2	Sequence 2, Appl
12	232.5	3.8	1089	1 US-08-177-431-2	Sequence 2, Appl
13	232.5	3.8	1089	5 PCT-US95-16930-2	Sequence 2, Appl
14	230.5	3.6	1664	1 US-09-599-652-2	Sequence 2, Appl
15	230.5	3.6	1664	1 US-08-642-846-2	Sequence 2, Appl
16	220.5	3.6	1664	4 US-09-264-604-2	Sequence 1, Appl
17	220.5	3.6	3248	1 US-08-353-700-1	Sequence 1, Appl
18	220.5	3.6	3248	5 PCT-US95-16216-1	Sequence 4820, Ap
19	219.5	3.6	1211	4 US-09-134-001C-4820	Sequence 28, Appl
20	219	3.6	1211	4 US-08-431-080-28	Sequence 28, Appl
21	219	3.6	1085	4 US-08-938-534-28	Sequence 28, Appl
22	219	3.6	1085	4 US-09-345-294-28	Sequence 94, Appl
23	216.5	3.6	1898	1 US-08-056-200-94	Sequence 94, Appl
24	216.5	3.5	2482	1 US-08-800-644-94	Sequence 6, Appl
25	212.5	3.5	1312	1 US-08-328-254-6	Sequence 29, Appl
26	212.5	3.5	1312	1 US-09-345-882-29	Sequence 4, Appl
27	206.5	3.4	764	1 US-08-375-300-4	Sequence 4, Appl

28	206.5	3.4	764	3 US-09-177-431-4	Sequence 4, Appl
29	206.5	3.4	764	5 PCT-US95-16930-2	Sequence 4, Appl
30	203.5	3.4	1162	2 US-08-728-323A-2	Sequence 2, Appl
31	203.5	3.4	1162	4 US-09-298-568-2	Sequence 2, Appl
32	203.5	3.4	1588	5 PCT-US93-07261-11	Sequence 11, Appl
33	203.5	3.4	1663	5 PCT-US93-07261-16	Sequence 16, Appl
34	203	3.3	1507	3 US-08-929-329-5	Sequence 1, Appl
35	202	3.3	1939	4 US-09-310-187A-1	Sequence 4, Appl
36	197	3.2	1177	4 US-09-134-001C-5106	Sequence 5106, Ap
37	196.5	3.2	1435	2 US-08-568-459A-4	Sequence 4, Appl
38	196.5	3.2	1435	2 US-08-487-826B-4	Sequence 4, Appl
39	196.5	3.2	1435	4 US-09-210-288-4	Sequence 148, Appl
40	193	3.2	1312	2 US-08-592-126-148	Sequence 51, Appl
41	193	3.2	1312	2 US-08-687-080-51	Sequence 2, Appl
42	193	3.2	1354	3 US-08-685-871-2	Sequence 5314, Ap
43	190.5	3.1	930	4 US-09-134-001C-5314	Sequence 3, Appl
44	190	3.1	1886	4 US-08-938-105-3	Sequence 4318, Ap
45	188.5	3.1	676	4 US-09-134-001C-4318	

ALIGNMENTS

RESULT 1
US-08-308-818-1
Sequence 1, Application US/08308818
Patent No. 5847077
GENERAL INFORMATION:
APPLICANT: Green, Michael R
TITLE OF INVENTION: A No. 5847077e1 Fungal Multisubunit Protein
TITLE OF INVENTION: Complex Critical for Expression of Fungal Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,818
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE/DOCKET NUMBER: 0342/0A404
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-52707700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1066 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE: S. cerevisiae
ORGANISM: S. cerevisiae
IMMEDIATE SOURCE: TAF-145
CLONE: US-08-308-818-1
Query Match 25.8%; Score 1569; DB 2; Length 1066;
Best local Similarity 32.3%; Pred. No. 6.6e-104;

Wed Mar 12 10:13:17 2003

us-09-601-965c-2.raii

Page 2

	Matches 390;	Conservative 204;	Mismatches 354;	Indels 260;	Gaps
QY	38	GQPLTNDDEWMAAQAASQDIALFG-----	NSNPDNSNI-----	SHDH 77	
Dd	8	GKTLMANADE-----	AYEALFGEGGSLSEISYIGDEGCANSKDTLELPDAVDEDE	62	
QY	78	MGGSNGCIIDNNHHSVNDHD-----	GLPNNLGNGNHLDDDDGCLN-----	DIGEL 124	
Dd	63	LADDDDLPEE-----	SDANLHPAMMTGAYDDVENGAVLGIDSNLNNQLEINCLDSIQ	119	
QY	125	F---D-----	QOESNVIN---TTHKHD-----	DDSN 148	
Dd	120	FILEDDGCFATSNALPFGMLANEIHLATETGVLDDSGANETIGSOLSIGVCNGNDMSIN		179	
QY	149	DGTLAEDDQREKENRRLQCKLOKLYKHLEKEDIKRINKIKYFPFYSRHREPFHK--FFS		207	
Dd	180	GGFIMPEPDMSDGKHKKATCL-----	LNH-EKYLKK-----	YFDFEFGKLKKKKKILYR 230	
QY	208	PSFOYRYQAPALASLNKIKELIPRTVKNLEIEMDOKAIFILRSADTASLSHEDKNVTNT		267	
Dd	231	RSVPYHMHSE-----	ISRVKPFMPPLMKLFKVQDDKRLFLMSRITSIYAPYIOGNN--NL	284	
QY	268	QDDLDPIKMLSEKRSRSDISFKEIDVYKRWTCQDFH-----	YSKDLVLTDDDDAI	323	
Dd	285	QSN-----	SSASRGLIHVSDILPFIKIQOKR-KRIIHDKETISEDLATATDDQEKI	338	
QY	324	INAGNEVYIVKPINELNLPNPLDNKQNRKLENDNTNTNNQNNNSVODEED-----		378	
Dd	339	INOGTSSTATL-----	ADSSMTPLNLF--SGGI-KLKSLEDAVEDAQWDE	381	
QY	379	DIPIFGQIILDK-LKIDMDPRLLPVPSK-----	KVADKATSVPESTDKLLEKFN	428	
Dd	382	DMIIDAKLKESKHAELENMDERLELLMIETNNLAAQOQDSSNLLPLNETLLQOKFN		441	
QY	429	SNDQVEYLLRKRYNTKORSQNLNLEIHSVPALRYOTPTKYKSLSTDETRSPHRYF--N		486	
Dd	442	SNDKQYQILKTKHQKRVSTI:NLNLIHQSPAINQDSPFYKVAVPYQLRHPRHREFGSH		501	
QY	487	VRPQTLVSEFSKILKRRKKDKGSLQOIFESKSLQTVADTGMIALEYSEQYRPILSNFG		546	
Dd	502	IRPQTKIYFSKLKARRRRKQDKGVKESFSTQDITIGDIAVYLMKESQCPVALSKRG		561	
QY	547	MGSGLINLYNRKERNPDRPFAOIGETHNLIYEDSPPMNGEVAAPGDVPLPILYNNVRA		606	
Dd	562	MANKLINLYNRKANEODTLRPKLPGETHVLGVQOKSPRMNGEVPFGHILPPLLYNNMTRA		621	
QY	607	PIFKHDKPFDLLVYKSGAGSHQKPYRGJNFPNVAVENTPY--EVPAPHSKVTATISNK		665	
Dd	622	PVEKHNDISGDFLLTKSSGFGISNRFYLRNINHLFTVGQTPVEEILPBPNSKRYTSMAKT		681	
QY	666	RLKVVYRVNNSLQVPIISVYDVKHNPENSDMNRQKLEFMYQORGEQOQYKWKVGL		725	
Dd	682	RLKMIYIRLINHNSKAISIDRIKHPRPDDQYGNQKVKFEMKYQYHDPKGLIKRMLD-		740	
QY	726	NDVIPEEELRTMTIPDDSSLYMTCMGQOVALDNDNVLFGDSQMSQESRSKKGDKREDST		785	
Dd	741	DEKLLINEAVKSLITPEOISQVZSMGOLFQEDN-----		775	
QY	786	ADDAENGDDINKDKEREVEKEQREEREKQDKREKDKKOTEKEKSKSKSEQDTEI		845	
Dd	776	--EAYNFD-----	SKLKS-----	786	
QY	846	DVEEELAPNLSKRFVIANOTK:MLDNGEGDPTGTGFGSMLRATQKNNPKPLTPRPE		905	
Dd	787	-LEENLPLWNITKFNINSTOMPRAMIOIHGVDPTGGEGEFSFKTSKGGFVYSGSSSN		845	
QY	906	NVPK-----	SNAAHNQKLYEDELKRYMWSORSLSYDHDGCTESKLAQOITNEYPRAD	957	
Dd	846	NNSSKKGTNTSHYNAOQKAVDEIATWTYHTHS-----		882	
QY	958	HELULKNKLEDOQOYQOQOQDPDLADQOQOQOQOQOQOQNRVLKTRRVRDENGIVHRVEFT		1017	
Dd	883	--LSISNPF-----	EMTPDEINQTNKKNVKTDRDQKTLKLYKRRKDENGIILQRPOTFI	935	

```

QY      1018 HDPELLIAAYVARKKIOEDDELLKNAVDVEIILPTNPKLEKTRRKLEKIANLEKRAKQSR 1077
Db      936 RDRFVIGIGYIKIRKIOEDKEDVKNKLEEDTSKINNELELEK QKRLLOJELANLEKSGQRR 994
QY      1078 AKRPFKLIHAAAAAGGATITDANTVMLPDGSYVYGCGCI-CKGSKPRRCKNGCAVGHIR 1136
Db      995 ARONSK-----RNGGATPRTENVSDNGSDLAGVTDCGAANKKNTTRCATCGQIGHIR 1048
QY      1137 TNAKCPY 1144
Db      1049 TNKSCPMY 1056

RESULT 2
US-08-188-582-14
; Sequence 14, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O. J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1872 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-188-582-14

Query Match 12.6%; Score 765.5; DB 1; Length 1872;
Best Local Similarity 22.5%; Pred No. 46-46;
Matches 314; Conservative 218; Mismatches 435; Indels 427; Gaps 54;

QY      43 TNDSEMAAOAAAESGFDALFGNSND-----PDSNLSHH----- 75
Db      24 TDSDEDSAGCGFSLAGLFLFGNINAGOLEGSESVYDDECKKRLAGALGAGLSLITETAYTA 83
QY      76 -DHMGDSNGIIRDN-----HHSSVND--HD-----GLFNNLNGNHHLLDD 113

```